

Primijenjena i inženjerska matematika

PRVI KOLOKVIJ

ZADATAK 1: [80 bodova]

Baza podataka tlak sadrži podatke o krvnom tlaku za ispitanike jedne ankete:

- varijabla spol je spol ispitanika,
 - varijabla dob sadrži broj godina za svakog ispitanika,
 - varijable sistolički-tlak i dijastolički-tlak sadrže vrijednosti sistoličkog i dijastoličkog tlaka,
 - varijabla tlak klasificira vrijednosti sistoličkog i dijastoličkog tlaka u tri kategorije: N - nizak tlak, O - normalan tlak, P - povišen tlak,
 - varijabla puls sadrži broj otkucaja srca u minuti (puls) za svakog ispitanika,
 - varijabla opće-stanje sadrži subjektivnu ocjenu (u standardnoj skali od 1 do 5) vlastitog zdravstvenog stanja.
- (a) [10 bodova] Za svaku varijablu odredite kojeg je tipa.
- (b) [10 bodova] Odredite tablicu frekvencija i relativnih frekvencija za varijablu tlak. Koliki je udio ispitanika s povišenim tlakom?
- (c) [10 bodova] Izračunajte aritmetičku sredinu, varijancu, standardnu devijaciju, medijan, donji i gornji kvartil te mod za varijablu sistolički-tlak. Koji je najveći izmjereni tlak?
- (d) [10 bodova] Nacrtajte kutijasti dijagram za varijablu sistolički-tlak u ovisnosti o spolu. Što zaključujete?
- (e) [10 bodova] Procijenite vjerojatnost da je sistolički tlak veći od 125 i vjerojatnost da je između 110 i 130.
- (f) [10 bodova] Intervalom pouzdanosti 95% procijenite očekivani sistolički tlak.
- (g) [10 bodova] Je li varijabla sistolički-tlak i normalno distribuirana na razini značajnosti 0.05?
- (h) [10 bodova] Odredite zajedničku empirijsku distribuciju varijabli spol i tlak. Jesu li te dvije varijable nezavisne na razini značajnosti 0.05?

ZADATAK 2: [20 bodova]

Varijable $placa_{2007}$ i $troškovi_{2007}$ u bazi podataka `gradjevina.sta` sadrže podatke o mjesečnoj plaći zaposlenika i ukupnim troškovima u 2007. godini za 100 građevinskih poduzeća srednje veličine u nekoj zemlji.

- (a) [10 bodova] Skicirajte dijagram raspršenosti te zaključite ima li smisla linearnom vezom modelirati varijablu $troškovi_{2007}$ (zavisna varijabla) u ovisnosti o varijabli $placa_{2007}$ (nezavisna varijabla). Možemo li na razini značajnosti 0.05 tvrditi da su te varijable korelirane?
- (b) [10 bodova] Procijenite pripadni regresijski pravac i zapišite njegovu formulu. Prema modelu, za koliko se povećaju troškovi ako se plaća poveća za 1?

ZADATAK 1

a)

- spol – kvalitativna nominalna varijabla
- dob - numerička nerekidna varijabla
- sistolički tlak – numerička neprekidna varijabla
- dijastolički tlak – numerička neprekidna varijabla
- tlak – kvalitativna ordinalna varijabla
- puls – numerička diskretna varijabla
- opće stanje – kvalitativna ordinalna varijabla

b)

Frequency table: tlak (tlak.sta)				
Category	Count	Cumulative Count	Percent	Cumulative Percent
N	7	7	14,00000	14,0000
O	25	32	50,00000	64,0000
P	18	50	36,00000	100,0000
Missing	0	50	0,00000	100,0000

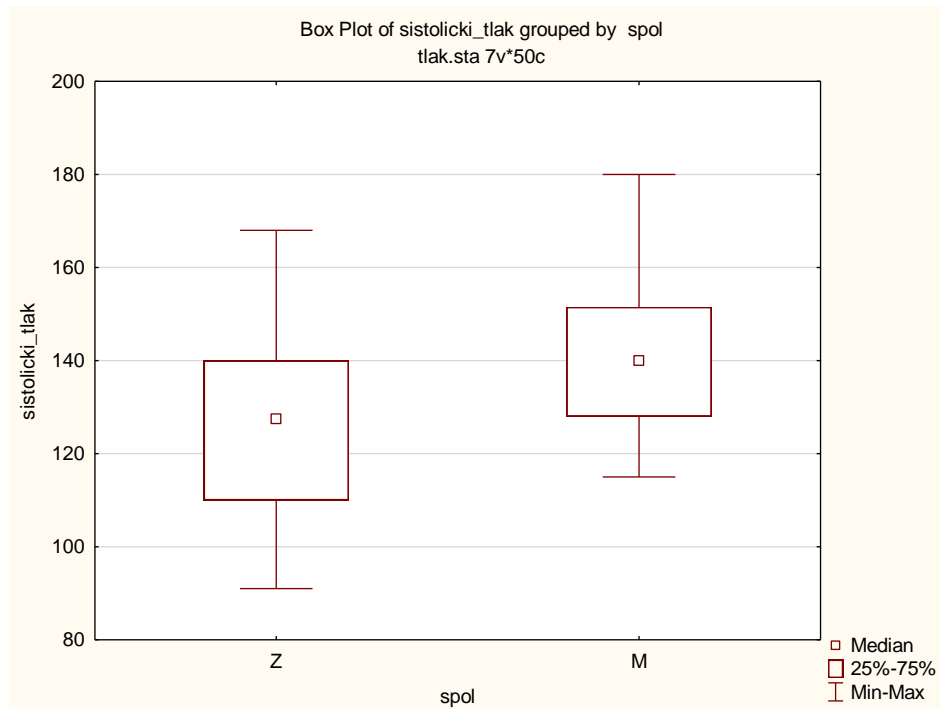
- Udio ispitanika s povišenim tlakom je 36 %.

c)

Variable	Descriptive Statistics (tlak.sta)										
	Valid N	Mean	Median	Mode	Frequency of Mode	Minimum	Maximum	Lower Quartile	Upper Quartile	Variance	Std.Dev.
sistolicki_tlak	50	133,7600	130,5000	140,0000	5	91,00000	180,0000	120,0000	146,0000	447,3290	21,15015

- Najveći izmjereni tlak je 180.

d)



- Medijan vrijednosti sistoličkog tlaka muškaraca veći je od medijana vrijednosti sistoličkog tlaka kod žena.
- Kod žena medijan sistoličkog tlaka iznosi 127.5 te je barem 50% izmjerenih vrijednosti sistoličkog tlaka manje ili jednako i barem 50% izmjerenih vrijednosti sistoličkog tlaka veće ili jednako od 127.5.
- Najmanja izmjerena vrijednost (minimum) sistoličkog tlaka iznosi 91, a najveća izmjerena vrijednost (maksimum) sistoličkog tlaka iznosi 168.
- Kod žena je barem 25% izmjerenih vrijednosti sistoličkog tlaka manje ili jednako 110 i barem 75% vrijednosti sistoličkog tlaka je veće ili jednako donjem kvartilu (110), dok gornji kvartil (140) govori da je barem 25% vrijednosti sistoličkog tlaka veće ili jednako 140 i barem 75% manje ili jednako od te vrijednosti.
- Kod muškaraca medijan sistoličkog tlaka iznosi 140 te je barem 50% izmjerenih vrijednosti sistoličkog tlaka manje ili jednako i barem 50% izmjerenih vrijednosti sistoličkog tlaka veće ili jednako od te vrijednosti.
- Najmanja izmjerena vrijednost (minimum) sistoličkog tlaka iznosi 115, a najveća izmjerena vrijednost (maksimum) sistoličkog tlaka iznosi 180.
- Kod muškaraca je barem 25% izmjerenih vrijednosti sistoličkog tlaka manje ili jednako 128 i barem 75% veće ili jednako donjem kvartilu (128), dok gornji kvartil (151.5) govori da je barem 25% vrijednosti veće ili jednako 140 i barem 75% manje ili jednako od te vrijednosti.

e)

Category	Frequency table: sistolički_tlak (tlak.sta)			
	Count	Cumulative Count	Percent	Cumulative Percent
91	1	1	2,00000	2,0000
100	2	3	4,00000	6,0000
105	3	6	6,00000	12,0000
110	3	9	6,00000	18,0000
115	1	10	2,00000	20,0000
116	1	11	2,00000	22,0000
117	1	12	2,00000	24,0000
120	2	14	4,00000	28,0000
122	1	15	2,00000	30,0000
125	1	16	2,00000	32,0000
126	2	18	4,00000	36,0000
127	2	20	4,00000	40,0000
128	2	22	4,00000	44,0000
130	3	25	6,00000	50,0000
131	2	27	4,00000	54,0000
138	2	29	4,00000	58,0000
139	1	30	2,00000	60,0000
140	5	35	10,00000	70,0000
142	1	36	2,00000	72,0000
145	1	37	2,00000	74,0000
146	1	38	2,00000	76,0000
149	1	39	2,00000	78,0000
152	1	40	2,00000	80,0000
154	1	41	2,00000	82,0000
155	1	42	2,00000	84,0000
159	1	43	2,00000	86,0000
164	1	44	2,00000	88,0000
165	1	45	2,00000	90,0000
166	1	46	2,00000	92,0000
168	1	47	2,00000	94,0000
170	1	48	2,00000	96,0000
173	1	49	2,00000	98,0000
180	1	50	2,00000	100,0000
Missing	0	50	0,00000	100,0000

- X - sistolički tlak
- Procjena vjerojatnosti da je sistolički tlak veći od 125: $P(X > 125) = 68\% = 0.68$
- Procjena vjerojatnosti da je sistolički tlak između 110 i 130: $P(110 < X < 130) = 26\% = 0.26$

(priznaje se i procjena vjerojatnosti $P(110 \leq X \leq 130) = 26\% = 0.38$)

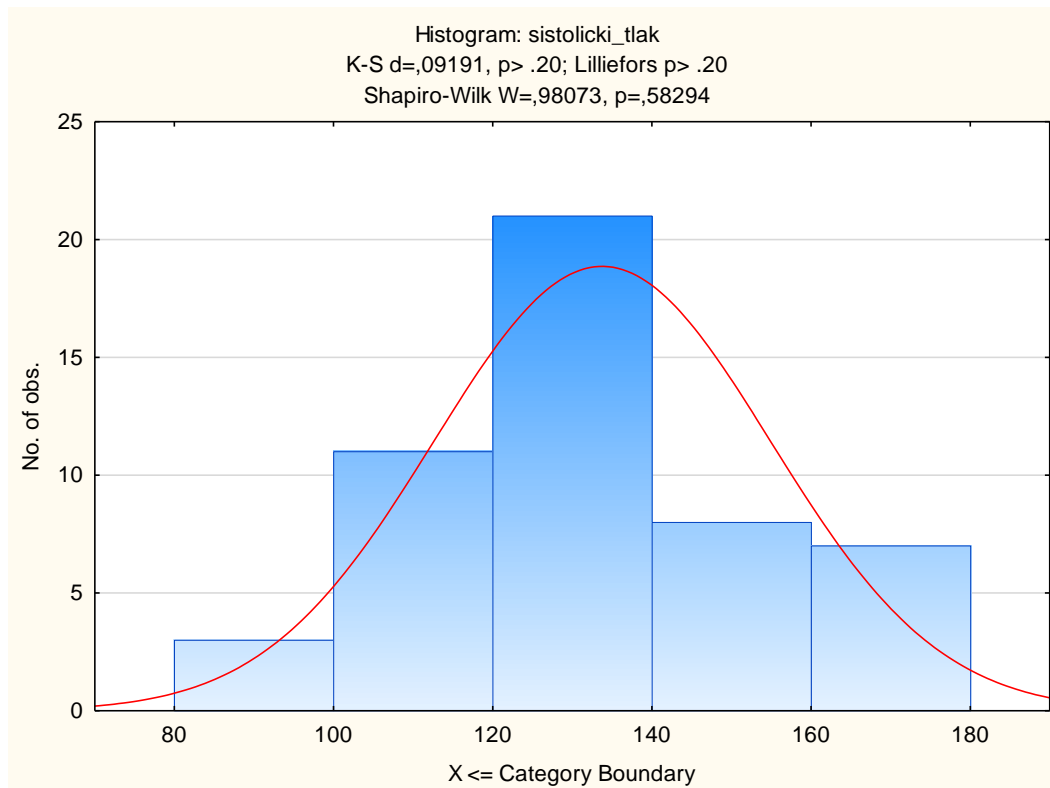
f)

Variable	Descriptive Statistics (tlak.sta)						
	Valid N	Mean	Confidence -95,000%	Confidence 95,000%	Minimum	Maximum	Std.Dev.
sistolicki_tlak	50	133,7600	127,7492	139,7708	91,00000	180,0000	21,15015

- Interval pouzdanosti 95% je slučajni interval koji s vjerojatnošću 0.95 sadrži stvarnu, nama nepoznatu, vrijednost očekivanja, a na temelju ovih podataka realizira se intervalom realnih brojeva [127.7492, 139.7708].

g)

- H_0 : sistolički tlak je normalno distribuiran
- H_1 : sistolički tlak nije normalno distribuiran
- α (razina značajnosti) = 0.05



- $p = 0.58294$ (Shapiro-Wilk); $p > 0.2$ (K-S test) $\rightarrow p > 0.05$
- Na razini značajnosti α (0.05) ne odbacujemo hipotezu H_0 tj. ne možemo tvrditi da sistolički tlak nije normalno distribuiran.

h)

Summary Frequency Table (tlak.sta) Marked cells have counts > 10 (Marginal summaries are not marked)					
	spol	tlak N	tlak O	tlak P	Row Totals
Count	Z	6	15	9	30
Total Percent		12,00%	30,00%	18,00%	60,00%
Count	M	1	10	9	20
Total Percent		2,00%	20,00%	18,00%	40,00%
Count	All Grps	7	25	18	50
Total Percent		14,00%	50,00%	36,00%	

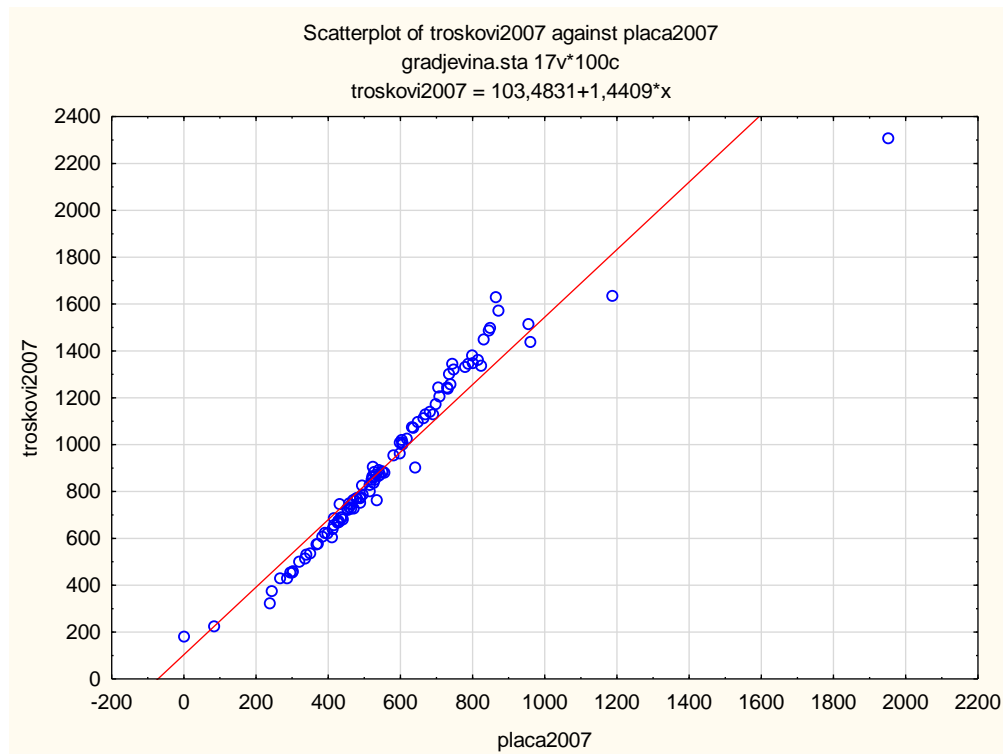
2-Way Summary Table: Observed Frequencies (tlak.sta) Marked cells have counts > 10				
spol	tlak N	tlak O	tlak P	Row Totals
Z	6	15	9	30
Total %	12,00%	30,00%	18,00%	60,00%
M	1	10	9	20
Total %	2,00%	20,00%	18,00%	40,00%
Totals	7	25	18	50
Total %	14,00%	50,00%	36,00%	100,00%

Statistics: spol(2) x tlak(3) (tlak.sta)			
Statistic	Chi-square	df	p
Pearson Chi-square	2,678571	df=2	p=,26203
M-L Chi-square	2,955656	df=2	p=,22813

- H_0 : varijabla spol i tlak su nezavisne
- H_1 : varijabla spol i tlak nisu nezavisne
- $\alpha = 0.05$
- $p = 0.26203$ (Pearson Chi-square); $p = 0.22813$ (M-L Chi-square) $\rightarrow p > \alpha$
- Na razini značajnosti α (0.05) ne odbacujemo H_0 , tj. ne možemo tvrditi da varijabla spol i tlak nisu nezavisne.

ZADATAK 2

a)



- Ima smisla linearnom vezom modelirati varijablu troškovi2007 u ovisnosti o varijabli placa2007.
- Koreliranost:
 - H_0 : varijable nisu korelirane
 - H_1 : varijable su korelirane
 - $\alpha = 0.05$

Spearman Rank Order Correlations (gradjevina.sta) MD pairwise deleted Marked correlations are significant at $p < ,05000$		
Variable	placa2007	troskovi2007
placa2007	1,000000	0,993447
troskovi2007	0,993447	1,000000

- $p < 0.05 \rightarrow p < \alpha$
- Na razini značajnosti odbacujemo H_0 i prihvaćamo H_1 , tj. možemo tvrditi da su varijable troškovi 2007 i plaća 2007 korelirane.

b) x – plaća (nezavisna varijabla); y – troškovi (zavisna varijabla)

- Regresijski pravac: $y = 103.4831 + 1.4409 \cdot x$
- Ako se plaća poveća za 1, troškovi će se povećati za 1.4409.